Abstract

The three parameter Dagum distribution is considered. The well known classical method – maximum likelihood (ML) estimation of three parameters is used which resulted in iterative solutions of the parameters. Hence, we propose a more efficient and computationally simple/analytical estimation procedure called two-step estimation. Using real data sets the procedure illustrated.
1. Introduction

Generally the lifetime data used for statistical analysis follows a particular statistical distribution. With knowledge of the appropriate distribution of any phenomenon we can greatly improve the sensitivity, power and efficiency of the statistical tests associated with it. Several distributions exist for modelling these lifetime data, however, some of the lifetime data do not follow the existing distributions or are inappropriately described. This therefore creates room for developing new distributions which could better describe some of the phenomena and provide greater flexibility in the modelling of lifetime data. One such distribution is the Dagum distribution.

Dagum distribution was introduced by Dagum [1] for modelling personal income data as an alternative to the Pareto and log-normal distributions. This distribution has been extensively used in various fields such as, income and wealth data, meteorological data, reliability and survival analysis. The Dagum distribution is also known as the inverse of Burr XII distribution. An important characteristic of Dagum distribution is that its hazard function decreases monotonically. This behaviour is described by several authors who have studied the model in different fields. In fact the Dagum distribution has been studied from a reliability point of view and is used to analyze the survival data. Domma et al. [2] estimated the parameters of Dagum distribution with censored samples, Kleiber and Kotz [4] provided an exhaustive review on the origin of Dagum model and its applications [5]. Srinivasa Rao and Kantam [7] have given a two-stage estimation for log logistic distribution. In this paper estimation of parameters of the Dagum model is studied. Though this is attempted by other researchers in the past by Kundu and Raqab [6], Ravikumar and Kantam [8]. Our approach of estimation is different from theirs.
The probability density function (pdf) of Dagum distribution is given by

$$f(x) = \frac{ap}{x} \left( \frac{x}{b} \right)^{ap} \left( \frac{x}{b} + 1 \right)^{-p-1}, \quad x > 0, \; a > 0, \; b > 0, \; p > 0. \quad (1)$$

Its cumulative distribution function (cdf) is

$$F(x) = \left( 1 + \left( \frac{x}{b} \right)^{-a} \right)^{-p}, \quad x > 0, \; a > 0, \; b > 0, \; p > 0. \quad (2)$$

The Dagum distribution is a skewed, unimodal distribution on the positive real line. The mean, median, and variance of the Dagum distribution are, respectively.

$$\text{Mean} = -\frac{b}{a} \frac{\Gamma \left( -\frac{1}{a} \right) \Gamma \left( \frac{1}{a} + p \right)}{\Gamma p} \quad \text{if} \; a > 1, \quad (3)$$

$$\text{Median} = b \left( -1 + 2^p \right)^{-\frac{1}{a}}, \quad (4)$$

$$\text{Variance} = \left\{ -\frac{b^2}{a^2} \left( \frac{\Gamma \left( -\frac{2}{a} \right) \Gamma \left( \frac{2}{a} + p \right)}{\Gamma p} + \left( \frac{\Gamma \left( -\frac{1}{a} \right) \Gamma \left( \frac{1}{a} + p \right)}{\Gamma p} \right)^2 \right\} \quad \text{if} \; a > 2. \quad (5)$$
2. Two-Step Estimation

The likelihood function is given by

\[
L = \prod_{i=1}^{n} f(x_i) = \frac{(ap)^n \left( \prod_{i=1}^{n} x_i^{(ap-1)} \right)}{(b)^{nap} \left( \prod_{i=1}^{n} \left( 1 + \left( \frac{x_i}{b} \right)^a \right)^{p+1} \right)} .
\]  

Log likelihood function of Dagum distribution is

\[
L(a, b, p) = n \log a + n \log p + (ap - 1) \sum_{i=1}^{n} x_i - nap \log b
\]

\[- (p + 1) \sum_{i=1}^{n} \log \left( 1 + \left( \frac{x_i}{b} \right)^a \right) .
\]

The score functions are given by

\[
\left\{ \begin{align*}
\frac{\partial LL}{\partial a} &= 0 \Rightarrow \frac{n}{a} + p \sum_{i=1}^{n} \log \left( \frac{x_i}{b} \right) = (p + 1) \sum_{i=1}^{n} \log \left( \frac{x_i}{b} \right) \\
\frac{\partial LL}{\partial p} &= 0 \Rightarrow \frac{n}{p} + a \sum_{i=1}^{n} \log \left( \frac{x_i}{b} \right) = \sum_{i=1}^{n} \log \left( 1 + \left( \frac{x_i}{b} \right)^a \right) \\
\frac{\partial LL}{\partial b} &= 0 \Rightarrow np = (p + 1) \sum_{i=1}^{n} \left( \frac{1}{1 + \left( \frac{x_i}{b} \right)^a} \right) .
\end{align*} \right.
\]

The maximum likelihood estimators are \( \hat{a}, \hat{b}, \) and \( \hat{p} \) are obtained by Equating (8) and solution for the non-linear system of equations is done iteratively. But to solve the above system of equations iteratively in with
the parameters $a$, $b$, $p$ is quite tedious. To overcome this difficulty, we propose a different method of estimation of $\hat{a}$, $\hat{b}$, and $\hat{p}$ based on the given sample. This procedure is called a two-step estimation as done by Engeman and Keefe [3]. In the present situation, we estimate $\hat{a}$, $\hat{b}$, and $\hat{p}$ as described below:

Consider the quantile functions of the Dagum distribution

$$Q(p) = b \left[ k\left(\frac{1}{p}\right)^{\frac{1}{a}} - 1 \right]^{\frac{1}{a}}.$$  \hspace{1cm} (9)

Let $X_1 < X_2$ be two ordered observations from the Dagum distribution. Let $q_1$ and $q_2$ be the corresponding sample taken as $\frac{1}{3}$ and $\frac{2}{3}$:

$$X_1 = b \left[ (q_1)^{\frac{1}{p}} - 1 \right]^{\frac{1}{a}},$$ \hspace{1cm} (10)

$$X_2 = b \left[ (q_2)^{\frac{1}{p}} - 1 \right]^{\frac{1}{a}},$$ \hspace{1cm} (11)

$$\frac{X_1}{X_2} = \frac{b \left[ (q_1)^{\frac{1}{p}} - 1 \right]^{\frac{1}{a}}}{b \left[ (q_2)^{\frac{1}{p}} - 1 \right]^{\frac{1}{a}}},$$ \hspace{1cm} (12)

$$\log\left(\frac{X_1}{X_2}\right) = \frac{1}{a} \log\left(\frac{(q_2)^{\frac{1}{p}} - 1}{(q_2)^{\frac{1}{p}} - 1}\right),$$ \hspace{1cm} (13)
\[
\hat{a} = \frac{\log \left( \frac{(q_2)^{-\frac{1}{p}} - 1}{(q_1)^{-\frac{1}{p}} - 1} \right)}{\log \left( \frac{X_1}{X_2} \right)}.
\]  

(14)

If a random sample of size \(n\) observations is given, it is ordered and divide into two halves called lower half and upper half. The arithmetic mean (A. M.) of each half would constitute a random ordered sample of size two. This can be used in Equation (14) to get \(\hat{a}\) using the entire sample. Accordingly \(\hat{b}\) and \(\hat{p}\) can be obtained with the help of Equations (8). Instead of arithmetic means of the lower and upper halves of a divided ordered sample, we can consider the geometric means (G. Ms) and Harmonic means (H. Ms) also to get estimators of \(\hat{a}\), \(\hat{b}\), and \(\hat{p}\). Thus a given random sample of size \(n\) generates 3 sets of estimates of the parameters \(\hat{a}\), \(\hat{b}\), and \(\hat{p}\), respectively using A.M., G.M., and H.M. of divided ordered sample. This method of estimating \(\hat{a}\), \(\hat{b}\), and \(\hat{p}\) is illustrated by an example in Section 3.

3. Application

3.1. Example 1

The newly proposed distribution was used to model the exceedances of flood peaks (in m³/s) of Wheaton River near Carcross in Yukon Territory, Canada. The data consists of 72 exceedances for the year 1958-1984, rounded to one decimal places given below:

1.7, 2.2, 14.4, 1.1, 0.4, 20.6, 5.3, 0.7, 1.4, 18.7, 8.5, 25.5, 11.6, 14.1, 22.1, 1.1, 0.6, 2.2, 39.0, 0.3, 15.0, 11.0, 7.3, 22.9, 0.9, 1.7, 7.0, 20.1, 0.4, 2.8, 14.1, 9.9, 5.6, 30.8, 13.3, 4.2, 25.5, 3.4, 11.9, 21.5, 1.5, 2.5, 27.4, 1.0, 27.1, 20.2, 16.8, 5.3, 1.9, 10.4, 13.0, 10.7, 12.0, 30.0, 9.3, 3.6, 2.5, 27.6, 14.4, 36.4, 1.7, 2.7, 37.6, 64.0, 1.7, 9.7, 0.1, 27.5, 1.1, 2.5, 0.6, 27.0.
In order to apply this example for our proposed distribution we have to confirm the goodness of fit of the Dagum distribution to the data in the Example 1. This is done through the well known QQ-Plot method, and the value we get is \( R = 0.9423 \). Hence, we conclude that the Dagum distribution provides a reasonable goodness of fit for the data.

**Table 1. Parameter estimation by two step method**

<table>
<thead>
<tr>
<th></th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( \hat{a} )</th>
<th>( \hat{b} )</th>
<th>( \hat{p} )</th>
<th>(-2\text{LL})</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.M.</td>
<td>2.8278</td>
<td>21.3278</td>
<td>0.5182</td>
<td>5.3347</td>
<td>3.1056</td>
<td>422.817</td>
</tr>
<tr>
<td>G.M.</td>
<td>0.9644</td>
<td>24.9159</td>
<td>0.3119</td>
<td>5.4294</td>
<td>3.1928</td>
<td>299.343</td>
</tr>
<tr>
<td>H.M.</td>
<td>0.1979</td>
<td>16.8467</td>
<td>0.4264</td>
<td>5.3071</td>
<td>3.2851</td>
<td>282.849</td>
</tr>
</tbody>
</table>

The lowest value of \(-2\text{LL}\) provides a performance of the pair of estimators over the other pairs. For Example 1, the preference goes to the pair obtained by using harmonic mean (H.M.) of the lower and upper halves of the divided ordered sample.

**Example 2.** Tensile strength of 100 observations of carbon fibers are

3.70, 3.11, 4.42, 3.28, 3.75, 2.96, 3.39, 3.31, 3.15, 2.81, 1.41, 2.76, 3.19, 1.59, 2.17, 3.51, 1.84, 1.61, 1.57, 1.89, 2.74, 3.27, 2.41, 3.09, 2.43, 2.53, 2.81, 3.31, 2.35, 2.77, 2.68, 4.91, 1.57, 2.00, 1.17, 2.17, 0.39, 2.79, 1.08, 2.88, 2.73, 2.87, 3.19, 1.87, 2.95, 2.67, 4.20, 2.85, 2.55, 2.17, 2.97, 3.68, 0.81, 1.22, 5.08, 1.69, 3.68, 4.70, 2.03, 2.82, 2.50, 1.47, 3.22, 3.15, 2.97, 2.93, 3.33, 2.56, 2.59, 2.83, 1.36, 1.84, 5.56, 1.12, 2.48, 1.25, 2.48, 2.03, 1.61, 2.05, 3.60, 3.11, 1.69, 4.90, 3.39, 3.22, 2.55, 3.56, 2.38, 1.92, 0.98, 1.59, 1.73, 1.71, 1.18, 4.38, 0.85, 1.80, 2.12, 3.65.

To confirm the goodness of fit of the Dagum distribution to the data in this example, the QQ-Plot technique is applied and the resultant value is \( R = 0.7777 \). Hence, we conclude that the Dagum distribution provides a reasonable goodness of fit for the data.
Table 2. Parameter estimation by two step method

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$\hat{a}$</th>
<th>$\hat{b}$</th>
<th>$\hat{p}$</th>
<th>$-2\text{LL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.M.</td>
<td>1.8206</td>
<td>3.4022</td>
<td>2.2172</td>
<td>2.9403</td>
<td>2.9403</td>
<td>183.045</td>
</tr>
<tr>
<td>G.M.</td>
<td>1.0205</td>
<td>3.8602</td>
<td>1.0419</td>
<td>3.2519</td>
<td>3.2519</td>
<td>170.625</td>
</tr>
<tr>
<td>H.M.</td>
<td>0.6805</td>
<td>3.6169</td>
<td>0.8299</td>
<td>3.2994</td>
<td>3.2994</td>
<td>165.745</td>
</tr>
</tbody>
</table>

The lowest value of $-2\text{LL}$ provides a performance of the pair of estimators over the other pairs. For this example, the preference goes to the pair obtained by using harmonic mean (H.M.) of the lower and upper halves of the divided ordered sample.

4. Summary and Conclusion

To solve the system of equations obtained in the maximum likelihood estimation iteratively are quiet tedious. To overcome this difficulty, we suggest a different method of estimation known as a two-step estimation to find the parametric values of the distribution.

References


DOI: https://doi.org/10.1080/02664763.2011.578613


DOI: https://doi.org/10.1016/0010-4809(85)90017-5


DOI: https://doi.org/10.1002/0471457175


DOI: http://dx.doi.org/10.18187/pjsor.v11i1.863

DOI: https://doi.org/10.1016/j.csda.2004.05.008
